How to Get There from Here:

scripted

Volume 5, Issue 2, August 2008

How to Get There from Here: Re-Use of Administrative Records in the Netherlands and the UK

Catherine Heeney[*]

Abstract

The traditional methods of record management and the legal system in the UK have not easily supported access to and reuse of public sector administrative records by National Statistical Institutes (NSIs). This paper will consider the measures being taken to address this situation. These include the Statistics and Registration Service Act 2007, which comes into effect in 2008 and strategies, such as the National Programme for IT (NPfIT), the aim of which is to standardise record management within the National Health Service (NHS). It will be argued that albeit that these are important facilitators of reuse of administrative records for research purposes, the ethical dimensions of such moves must be carefully thought through. This is important if the widespread social acceptance necessary for sustainable access and reuse of administrative records by NSIs is to be achieved. This paper will discuss how lessons can be learnt from the approach taken by the Netherlands to aspects of data management including their approach to access, reuse and future dissemination of statistical products.

DOI: 10.2966/scrip.050208.294

[*] Research Fellow, The Ethox Centre, Dept of Public Health and Primary Care, University of Oxford.

1. Introduction

This paper will consider the spurs and barriers for access and reuse of public sector administrative records by National Statistical Institutes (NSIs). NSIs are the principal public sector organisation involved in collecting, compiling and disseminating official statistics[1] within a nation state. The Dutch NSI, the Centraal Bureau voor De Statistiek (CBS) and the UK's NSI, the Office for National Statistics (ONS)[2] and their access and reuse of administrative data will be considered. CBS has enjoyed both access and reuse of public sector administrative data for several decades for the purposes of statistical research. ONS has for some time been seeking to establish comparable access to and reuse of administrative records collected by government bodies.[3] However, in the UK an ambiguous legal framework and a non-standardised system of records management have presented major barriers to the realisation of this ambition. The Netherlands has well established legal and technical systems to support record reuse. This has created a framework that has been of great importance in enabling CBS to make use of administrative records. However, it will be argued that acknowledgement of the social and ethical dimensions of this process have been vital to sustainable access and reuse of administrative data in the Netherlands. To illustrate the dangers of neglecting these dimensions, the paper will refer to the example of an unsuccessful attempt to establish long term data-linking in an administrative department of the Canadian government.

In the UK, the *Statistics and Registration Service Act 2007*, which came into effect in spring 2008, is among changes which aim at improving the exploitation of administrative records. The Act potentially heralds a new era of availability of administrative records for the production of statistics. It also provides an opportunity to consider some of the challenges to the reuse of administrative data and how far the new legislation can address them. Alongside legal developments, programmes such as the National Programme for IT (NPfIT) aim to address the traditional lack of standardisation in NHS records systems. The argument presented here is that while legislation and standardisation are important facilitators of the reuse of administrative records, the ethical and social impact of these measures must be addressed. Widespread social acceptance is a vital element of sustainable use of administrative records. The paper will discuss primarily how lessons can be learnt from the approach taken by the Netherlands with regard to the access, use and subsequent dissemination of statistical products to the wider research community. The following factors will be discussed in relation establishing sustainable access by public researchers to administrative data held by public bodies: clarity and transparency in regard to the types and purposes of access, and strict delimitation of subsequent use

---

[1] Official statistics disseminated by the national statistical institute if the system is centralised but may be released by other departments if the system is decentralised. See Organisation for Economic Cooperation and Development, Glossary of Statistical Terms, Available at http://stats.oecd.org/glossary/detail.asp?ID=4350.

[2] Soon to be the Statistics Board.

[3] D Wroe, "Beyond 2001: Alternative to the Census: Study for the Office for National Statistics" (1998) Volume 1 *Office for National Statistics.*

and dissemination of statistical products. It will be argued that these factors are important in ensuring the trust and support of the public and of the departments supplying the data, and that maintaining this support must be taken as seriously as the technical or practical, and legal aspects of data management.

There is increasing endorsement across the research community for access and reuse of records which have already been collected by the public sector.[4] As data is seen increasingly as a resource to be used beyond its original purposes, the paper also aims to address generic issues arising from the reuse of data outside of the context in which it was collected. The article will deal with the benefits both of access and reuse of data. Access here refers to the availability of data to a particular organisation. Reuse, in this context, means that the data is available for a purpose other than that for which it was collected and/or by parties other than the original collectors. In the case under discussion here, reuse of administrative records entails that administrative data collected by an administrative department is available for use by NSIs for statistical activities. The desirability of such an approach has been seen as advantageous for a considerable time. In 1979 a United Nations report on the subject stated, that "Administrative data may be an inexpensive source of data for the statistician, because they arise as a by-product of ongoing activities paid for out of other budgets."[5] The CBS in the Netherlands has access to administrative records for statistical research and this is advantageous for a variety of reasons. For example, one of the main reasons cited is that reuse of administrative data is cost-effective. The UK Census is a costly exercise. It requires an army of people to deliver forms to every household, follow up exercises to capture slow responders and a substantial advertising budget.[6]

2. Official Statistics in the UK

In the UK, the Census remains the most important source of statistical data on the whole population. The possibility of extending the use of administrative records by ONS has been under discussion for some time. If this were achieved it could reduce the number of costly surveys and augment or replace data collected in the Census.[7] Quite apart from the cost of carrying out a census, there are a number of problems concerning the Census data itself. The Census, while it constitutes the main source of statistical data on the whole UK population, is only taken every ten years. This means that the data will be relatively out of date for some periods of time.[8] Data from administrative sources is likely to be up to date and could potentially avoid the problem of non-response among target populations. Despite its being a legal

---

[4] The Wellcome Trust, "Sharing Data from Large-scale Biological Research Projects: A System of tripartite responsibility," *Report of a meeting organized by the Wellcome Trust* held on 14-15 January 2003 at Fort Lauderdale, USA.

[5] United Nations, "The Development of Integrated Data Bases for Social, Economic and Demographic Statistics" (1979) *United Nations*, New York.

[6] Between 2000 and 2002, £150 million was spent on the UK 2001 Census, See "Office for National Statistics, Outsourcing the 2001 Census" *Report by the comptroller and auditor general*, (2002) HC 1211 Session 2001-2002 18 October 2002, National Audit Office.

[7] D Wroe, *see* note 3.

[8] *Ibid.*

requirement to respond to the Census significant numbers of people still fail to participate. There was said to be undercount of one million people on the 2001 Census.[9] ONS and its predecessor organisations have traditionally had only limited and restricted use of administrative records.[10] For example, administrative data has been made available to ONS and its predecessors under the Population Statistics Act 1938. This required administrative data on life events to be provided to the Registrars General for statistical purposes. Administrative sources are also used for research in isolated and strictly limited cases such as the Longitudinal Survey, which links data from the records of some individuals to Census data.[11] The Census Act 1920 grants permission for a census to be carried out and was amended in 1991 by the Census Confidentiality Act but this legislation does not deal with access and reuse of administrative data.

The Office for National Statistics (ONS) in the UK was established in 1996 as an executive agency. It was a department of the Chancellor of the Exchequer. The Statistics and Registration Service Act 2007 transfers the "property, rights and liabilities" of ONS to the Statistics Board. [12] The Statistics Board is a department of the Cabinet Office. [13] The production of official statistics in the UK is decentralised. Statisticians who work on official statistics across government departments and in the devolved administrations of Wales, Scotland and Northern Ireland are part of an umbrella organisation, the Government Statistical Service (GSS). Statisticians working for the GSS produce National Statistics (NS). The Statistics Commission, which had responsibilities under the Code of Practice, was a non-statutory advisory body with responsibility for priority setting, the integrity of National Statistics and quality assurance.[14] This body will cease to exist when the SRSA comes into force in 2008. [15] The duties of this body will be subsumed by the Statistics Board.

In the absence of specific statistical legislation in the UK, the National Statistics Code of Practice and associated protocols were originally intended as an ethical and procedural guide for government statisticians. A National Statistics Code of Practice was produced in 2002 along with eight Principles and twelve Supporting Protocols in 2004. [16] The NS Code of Practice was intended to facilitate access to administrative records. This is evident from the description given of the roles of departmental ministers, which is to "authorise access to all data within their control for statistical

---

[9] N Cohen, "Our Missing Million" (2003) *The Observer*, 9 November 2003.

[10] The Population Statistics Act 1938 requires administrative data on life events to be provided to the Registrars General for statistical purposes.

[11] Data on vital events from administrative sources are matched to census data for a subset of the population in England and Wales.

[12] Statistics and Registration Service 2007 Section 1(1).

[13] HC/HL Deb. 25 Jul 2007, Cols 908-912, at:

http://www.publications.parliament.uk/pa/cm200607/cmhansrd/cm070702/debtext/70702-0009.htm

[14] National Statistics, "Framework for National Statistics", First Edition National Statistics, Section 4.1.7 (k) (2000).

[15] Part of its remit was to consider the appropriateness of specific statistical legislation and to make recommendations on this issue.

[16] National Statistics, "Code of Practice: Statement of Principle" at: http://www.statistics.gov.uk/about-ns/cop/downloads/StatementRD.pdf.

purposes across government, subject to confidentiality considerations and statutory requirements."[17] However, the existence of National Statistics did not in itself solve the problem for ONS of access to administrative data held by other departments. Government departments must act within the law and there had previously been no well defined legal "gateway" for administrative records held by government departments to be accessed by ONS for the production of statistics. [18]

3. The Production of Official Statistics in the Netherlands

The system of official statistics in the Netherlands is centralised, with the CBS producing the majority of government statistics. The CBS, as of 2004, is an autonomous agency but the Minister for Economic Affairs remains politically responsible for legislation and the budget.[19] The Social Fiscal Number, now the Burgerservicenummer (BSN) is a unique national identifier allocated to Dutch citizens at birth and used consistently in public sector records. The population register contains the BSN, which is extensively used in government administrative records. A unique national identifier is a number or code assigned to each individual and where this is used consistently in most records, as is the case in the Netherlands, it greatly facilitates the linkage of records held on individuals by different government departments or organisations. A population register is an index or list which contains a unique identifier and other information about an individual. This allows a record to be matched to an individual's unique number and verified. Used in conjunction the identifying number and the population register provide invaluable tools for successfully matching records on a given individual. This system facilitates the creation of a flexible dataset. Flexibility is necessary if administrative records are to provide data on the range of variables available from censuses and surveys.

The coverage achieved by the population register is said to be good due to its central role in the Dutch public sector. As Van der Laan points out "It is extremely difficult to function in Dutch society without being included in the population administration."[20] There is no Census in the Netherlands. Instead the creation of Census type data involves linking micro-data from administrative sources with household sample surveys, which are conducted by the CBS. The prevailing view at CBS is that a Dutch Census "[I]s too costly and uncertain to serve as an adequate source for policy-oriented purposes and, therefore it has become obsolete as a meaningful statistical data source."[21] Administrative records are seen as the primary source of information on the population.[22] The continued functioning of the current system in the Netherlands depends upon government departments being both able and

---

[17] National Statistics 2000, *See* note 14.

[18] P Jackson, "The Legal Framework for National Statistics in the UK" 2003, *Office for National Statistics*, Joint ECE/Eurostat work session on statistical data confidentiality, Luxemburg, 2003.

[19] United Nations Statistical Division, Country Profiles of Statistical Systems, United Nations, 2005 at: http://unstats.un.org/unsd/dnss/cp/searchcp.aspx.

[20] P Van der Laan, "Integrating administrative registers and household surveys" (2000) 15 *Netherlands Official Statistics,* Statistics Netherlands, 15.

[21] *Ibid* 17; Public objections to the 1971 Census were instrumental in the Dutch Parliament's decision to revoke the legal obligation to perform a population Census.

[22] *Ibid*.

willing to supply administrative records to CBS. This cooperation, moreover, rests upon these departments being reassured that the data they supply is being correctly managed and protected.[23] The Sociaal Statistisch Bestand (SSB) is a database designed to manage the linking of administrative data for statistical purposes.[24] The use of the SSB requires that identifiers are used for matching purposes. The security measures for SSB involve storing identifying data separately from data files, while the access to linked records is strictly controlled.

The *Dutch Personal Data Protection Act*, *Wet Bescherming Persoonsgegevens* (WBP), came into force in 2000 following EU Directive 95/46/EC. Article 9 of the WBP grants exemptions from the requirement to notify the Data Protection authority of data processing done for statistical purposes. Article 29 and 44 of the Act exempts the processor of statistical data from granting rights which would be accorded to data subjects in the case of "personal data." These include entitlement to informed consent and verification of the data. This exemption rests upon the condition that "the necessary arrangements have been made to ensure that the personal data can only be used for statistical or scientific purposes."[25] The production of official statistics was, regulated specifically by the Official Statistics Act 1996, which has now been superseded by the Statistics Netherlands Act 2003. The Statistics Netherlands Act establishes a legal basis for the existence of CBS, and the Central Commission of Statistics (CCS) which is a supervisory body. Section 3 of this Act defines the duty of CBS as being "to carry out statistical research for the government for practice, policy and research purposes and to publish the statistics compiled on the basis of such research." Sections 33 and 34 of the Act deal with access to administrative records for statistical purposes. Section 33 enables the use of registers from a wide variety of public sector institutions and departments (including central government, utility providers and local authorities) for statistical purposes. Section 34, permits the use of the BSN (formerly the Social Fiscal Number) for statistical purposes. The director general of CBS is authorised to use the number in registers and communicate with agencies using this number for the purposes of compiling statistics.

4. Dissemination

As the dissemination of data is perhaps the key role of NSIs it is useful to consider some aspects of the approaches taken in the UK and the Netherlands, particularly as they relate to the establishing and maintaining trust. CBS allows comparatively restricted access to micro-data (individual level data) for outside researchers. The Scientific Statistical Agency or *Wenshappelijle Statistisch Agentschapp* (WSA) is part of the Scientific Research Council of the Netherlands. Its role is to act as a liaison between the scientific community and CBS. It must balance the demands for greater access to micro-data with the imperative to maintain public trust. Detailed tables and

---

[23] P Kooiman, J R Nobel and L C Willenborg, "Data Protection at Statistics Netherlands" (1999) 14 *Netherlands Official Statistics*, special issue Disclosure Control, Statistics Netherlands, Voorburg/Heerlen, 1999.

[24] P Van der Laan, *see* note 20; The English name for this is the Social Statistical Database.

[25] Wet bescherming persoonsgegevens, Data Protection Act Netherlands English translation, available at: http://www.cbpweb.nl/en/structuur/en pag cbp.htm.

micro-data are important products of CBS.[26] CBS releases public use files but scientists who seek access to more detailed micro-data must fulfil certain criteria. These include affiliation to a university or an institution which is involved in pure policy or related research.[27] Researchers allowed access to this data are subject to a code of conduct. In this, it is stipulated that data will only be used for statistical purposes and results will be submitted to CBS for checks. This system is known as Micro-data Under Contract (MUC). On-site access is provided for data more disclosive data. Private companies are not entitled to CBS micro-data. Dissemination of data is, therefore, strictly controlled and organisations and purposes are constrained. Joris Nobel of CBS noted, with regard to the adequacy of these measures to fulfil the legal obligation to protect the data-subjects, "Beyond that, it seems that the *Registratiekamer* has no particular strong feelings about (our feelings about) our relations with respondents."[28]

Unlike in the Netherlands, in the UK there is not a single authority that acts as a liaison between the academic community and NS or ONS. There are arrangements with several academic bodies for specific ONS data products. A microdata release panel was instituted at ONS at the same time as the Code of Practice to make decisions on the suitability of microdata for release to third parties for research purposes. The ESRC Data Archive and The Centre for Census and Survey Research, based at the universities of Essex and Manchester respectively, also hold data sets and deal with requests for access from the research community. This means there has not been the same centralised control over access to statistical products as there exists in the Netherlands. This establishes a situation which is perhaps more open in terms of making existing data available to a wider variety of researchers including those working within the private sector. Some argue that this benefit of National Statistical products are maximised by extensive use. Such arguments may oppose a system which aims to protect privacy and ensure trust by employing strict access controls. The view held by some at CBS is that the fact that official statistics depend upon access to administrative records requires a rather stringent approach to both data release.[29] The Statistics and Registration Services Act does not challenge the existing ethos with regard to the more open dissemination of official statistical products in the UK. In Section 8 of the Act the duties of the new Statistics Board with respect to the monitoring of "quality", "good practice" and "comprehensiveness" in the "production and publication of official statistics" are established. Arguably these duties are important as they maintain the reputation of official statistics and those who produce them. However, some critics of the Act claim that it does not go far enough generally to ensure public trust.[30]

---

[26] L Willenborg and T deWaal, "Information Loss Through Global Recoding and Local Suppression" (1999) 14 *Netherlands Official Statistics*, special issue Disclosure Control, Statistics Netherlands, Voorburg/Heerlen.

[27] P Kooiman et al, *see* note 23.

[28] J Nobel, "Informed Consent: Buzzword or Panacea" *Proceedings of the ECE/Eurostat Conference,* Thessaloniki, March 1999, 1.

[29] P Kooiman et al, *see* note 23.

[30] T Holt, "The Statistics and Registration Service Act" (2007) 4(4) *Significance* 182–183.

5. Access to Administrative Records

Trust and privacy have long been recognised as important issues in relation to dissemination of statistical data.[31] They are equally crucial factors in sustaining the support of relevant parties for access and reuse of administrative and they must be acknowledged, as such, even efforts to overcome legal and technical barriers are perhaps the focus of attention. In the Netherlands, there has been a system of record storage and management, and legislation supporting the reuse of administrative data for official statistics in place for a number of decades.[32] There is also specific legislation defining both the responsibilities and rights of CBS with regard to access to and use and dissemination of publicly held administrative records. The legal position on matching data across departments is clear, thus removing a significant barrier to the use of administrative records.[33] This situation has enabled the use of a variety of public sector records for statistical research. Until recently the UK had neither the practical infrastructure nor enabling legislation to allow the widespread reuse of administrative records.[34] Where legislation does not give a clear position, access to the identifying information needed to reuse data for statistical purposes can be an uncomfortable issue for those involved.[35] The important differences between the two countries in relation to access and use of administrative records are not, however, limited to information management practices and legislation.[36] Moreover, changes in both areas may fail to produce the desired results where key actors, such as, for example, those who collect and manage data or those who provide it, withhold their support and cooperation. [37]

A lack of stakeholder support for data sharing practices can entirely undermine data-sharing exercises, such as access and reuse of administrative records for secondary purposes. The antecedents and consequences of this will be discussed in relation to the so termed Canada's 'Big Brother Database'. New practices of access and reuse of public sector data in the UK must also address the traditional lack of data-sharing of this sort. Due to there being little precedent for data-sharing for research purposes and other purposes in the UK there is likely to be a period of organisational and professional readjustment. This will involve both acceptance of and resistance to new standards and proposed practices by professionals and organisations. Moreover, the

---

[31] Office for National Statistics, *Statistics: A Matter of Trust,* (1998), Report on the Consultation Exercise 24 February – 31 May 1998, Office for National Statistics.

[32] *See* Statistics Netherlands, Special Issue, Integrating administrative registers and household surveys, Vol. 15, Voorburg/Heerlen, (2000).

[33] P Jackson, *see* note 18.

[34] Before the introduction of the Act some administrative data was made available to ONS and its predecessors for example the Population Statistics Act 1938 requires administrative data on life events to be provided to the Registrars General for statistical purposes.

[35] P Bellamy, C Raab, A Warren and C Heeney, "Institutional Shaping of Interagency, Working: Managing Tensions between Collaborative Working and Client Confidentiality" 17 (3) *Journal of Public Administration Research and Theory*, 405.

[36] P Bellamy et al, *see* note 35.

[37] S L Star & J R Griesemer, "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." (1989) 19(3) *Social Studies of Science*, 387-420.

implementation of new policies is unlikely to be a straightforward process. New practices will not simply replace old understandings, agreements and roles but will need to acknowledge and even build on them. As Timmermans and Berg claim "Standards will attempt to change and replace those practices, but […] the same standards need, to a certain degree, to incorporate and extend those routines." [38] Current measures, such as the creation of a "data spine," [39] which could play the role of a population register, are an example of the will to emulate countries like the Netherlands. However, there have been key actors in this area who have voiced doubts about the extent to which the Statistics and Registration Service Act 2007 will enable statisticians based in 'policy departments' to act autonomously.[40] The doubts centre on the failure of the Act to adequately address institutional constraints, which the statisticians will meet in practice.[41] This problem arises due the decentralisation of the production of official statistics in the UK, a situation which does not exist in the Netherlands. The UK and the Netherlands will continue to provide a valuable source of comparison as the new UK Statistics Act takes effect.

5.1 Removing Legal Barriers to Access and Re-Use in the UK

The *Statistics and Registration Service Act 2007*, gives a legal position on many issues which have long presented barriers to access for ONS and its predecessors. It defines the role of the new Statistics Board as being "to have the objective of promoting and safeguarding the production and publication of official statistics that serve the public good."[42] It deals, among other things, with definitions of "Official Statistics" and of questions of access to statistics by ministers. Section 39 deals with "confidentiality of personal information" and ss. 42 to 50 address the sharing of administrative data from public bodies with the Statistics Board. As in data-protection legislation in EU and members states, the Statistics and Registration Service Act emphasises disclosure control as a key component of the protection of research subjects.[43] Section 43 deals with information relating to NHS registration. It permits the Secretary of State to disclose patient registration information to the Statistics Board. This information includes, address details, date of birth and the patient identification number.[44] The uses of these identifying data are limited to the "production of population statistics."[45] Section 47 relates to the power to authorise disclosure to the Board. The Minister for the Cabinet Office may make regulations allowing a public authority to disclose information to the Board, where "the disclosure would otherwise be prohibited by a rule of law" or "the authority would not

---

[38] S Timmermans & M Berg "Standardization in action: Achieving local universality through medical protocols" (1997) 27(2) *Social Studies of Science,* 273-305 at 274.

[39] M Cross, "Keeping the NHS electronic spine on track" (2006) 332(7542) *British Medical Journal*, 659-8.

[40] T Holt, 2007, *see* note 30.

[41] *Ibid*.

[42] *Statistics and Registration Services Act 2007*, ss. 1(1) and 7(1).

[43] *Ibid*, s. 39(10).

[44] *Ibid*, s. 43(3a – c).

[45] *Ibid*, ss. 43(5) and 44 covers the same in relation to Wales.

otherwise have the power to make the disclosure."[46] A transition period from the National Statistics Code of Practice and its Principles and Protocols to the Statistics and Registration Services Act will begin with the Act's entry into UK law. The Statistics Board will draw up guidelines in the interim period.[47]

This legislation presents an important landmark in relation to access to and use of administrative records by the UK's NSI. However, the governance framework has not been the sole barrier to access and reuse of administrative records. The UK system has traditionally been one of isolated departmental records management systems. Personal identifiers have not been used consistently across government and public sector data sources and there is no population register. One study carried out on behalf of ONS found there had been as many as eighteen different formats for the NHS number.[48] This presents obstacles to standardisation. There have been a number of initiatives throughout government designed to facilitate the linking up and reuse of data. For example, ONS has been at the forefront of plans to create a system to enable the secondary use of administrative records for statistical research.[49] Within the NHS, the (NPfIT) aims to gain a more standardised records management system. The aims of this programme are to improve service delivery but could also support the use of health records for research.[50] Indeed, there is increasing recognition in a number of quarters that there is enormous potential research value in the records already collected and held by the NHS.[51] UK Biobank will for example access and use NHS administrative records for biomedical research.[52] However, the structure of government systems for the collection and storage of data in the UK do not support routine data linking. Furthermore, existing practices are unlikely to simply be superseded even if new technologies and new governance systems encourage sharing of data.

---

[46] *Ibid*, s. 47(1a and b).

[47] *Statistics and Registration Services Act 2007*, ss. 17 and 19.

[48] D Wroe, *see* note 3.

[49] The Office for National Statistics, "Proposal for an integrated population system" (2003) *Office for National Statistics*, October 2003, at:

http://www.statistics.gov.uk/downloads/theme_population/ipss.pdf. This issue has grown in importance as a variety of government departments and bodies have been encouraged to share information on everything from monitoring the vulnerable and potentially dangerous individuals (6 P et al, see note 35).

[50] M Cross, 2006, *see* note 39. Other parts of the research community are moving towards the building of large databases to allow the sharing of data by cancer-geneticists see Cancer Research UK, National cancer tissue bank now open for donations, Press Release, Wednesday 5 September 2007.

[51] Wellcome Trust, "Report on the Clinical Research Collaboration and Wellcome Trust Frontiers Meeting on the Use of Electronic Patient Records for Research and Health Benefit" held 24 and 25 May 2007, at:

http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd038686.pdf.

[52] Biobank, *Further Information Leaflet*, *see*

http://www.ukbiobank.ac.uk/docs/urtherinfoleaflet241007.pdf.

6. The Ethical Dimensions of Re-Use of Data

As argued above, aside from the technical and legal aspects of access to and reuse of administrative records, there are other important issues to be considered. These fall under the heading of the social and the ethical. They range from how the actors involved in the production of data incorporate both technology and law into practice, to matters of privacy and public good.[53] In the sphere of statistics the claim is often made by policy makers that the production of high quality statistics is a good thing precisely because they give rise to knowledge which will ultimately be beneficial, whether it be to people in nation states or more broadly.[54] The justification for the existence of National Statistical Institutes (NSIs) often rests on the argument that they provide the government with the knowledge to govern in a more enlightened and informed way.[55] These sorts of public good argument can also be made about the reuse of administrative data: using this data could save public money by providing the raw materials for statistical research. The counter argument usually comes from the perspective of informational privacy and individual rights to control information pertaining to them. It has been claimed that the issues are the largely the same for administrative data and for data originally collected for statistical purposes.[56] However, the concept of "contextual privacy", introduced by Nissenbaum, could undermine this claim.[57] If context is a key factor in an individual's decision to provide information, and Nissenbaum makes a convincing argument for this being the case, then the original context into which information is provided is not the same in the two cases. In other words, if one provides administrative data it is on the basis of accessing a service with its accompanying incentives and expectations. These will be quite different when one provides data for research or statistical purposes.

Otherwise, privacy is defined rather narrowly in the official governance framework for NSIs where it is seen simply as avoidance of the disclosure of identified information outside of the original collecting organisation. As indicated above, the law governing data more generally has exemptions for statistical and research data. Therefore, while the issue of context is recognised as being important in the protection of privacy for identifiable data or 'personal data'[58] it is not significant in the case of data which is to be used for research purposes.[59] The *Statistics and Registration Services Act* remains faithful to the *Data Protection Act* defining

---

[53] These can be seen as poles of a debate which is current in many spheres and is sometimes seen as a matter of balance between these two central values in a post-enlightenment society, whose value system can be broadly said to reflect liberal moral and political ideals.

[54] International Statistical Institute 1985, ISI Declaration on professional ethics, at: http://www.cbs.n1/isi/ethics.htm.

[55] Ibid.

[56] K Marquis, S Wetrogan and H Palacios, U.S. Census Bureau, "Towards a U.S. Population Data Base from Administrative Records" (1996) Proceedings of the Government Statistics Section, American Statistical Association, 117 -122.

[57] H Nissenbaum, "Protecting Privacy in an Information Age: The Problem of Privacy in Public" (1998) 17 *Law and Philosophy* 559-596.

[58] *Data Protection Act 1998*, available at http://www.hmso.gov.uk/acts/acts 1998/19980029.htm.

[59] W W Lowrance, "Learning for Experience: Privacy and Secondary Use of Data in Health Research" 2002 *The Nuffield Trust*.

"personal information" as "information which relates to and identifies a particular person (including a body corporate)." [60] In the area of statistical research protection of privacy tends to involve some combination of confidentiality and anonymity of data on the one hand, and informed consent on the other. However, arguably linking and reusing data challenges this traditional schema. In order to link records the possibility of tracing individuals to whom the data belongs must exist and, therefore, the data must remain identifiable in some manner. Where there is a population register which contains an index number this can act as a pseudonym or a proxy for identity. This is the system used by CBS when linking administrative files for statistical purposes. At some level, the links between the identity of data subjects and large amounts of data must be retained. The case of the database created by Human Resources Development Canada (HRDC), to be discussed below, illustrates how that failure to acknowledge and deal with these matters can cause concrete problems of purpose drift and lead to a widespread perception that public trust and individual privacy has been abused.

7. Lessons on Trust and Re-Use of Publicly Held Data

The lack of complaints from the Dutch public about the linking-up of administrative records for statistical purposes may rest on ignorance of these practices. However, more optimistically, it may be due to there being a sufficient level of trust in the ability of CBS to respect privacy while carrying out its activities. This is important as without some level of support from the public and policy bodies, efforts to link data may raise objections on privacy grounds and undermine the exercise. A useful reminder of the importance of trust in achieving public support comes from Canada. The case of the Human Resources Development Canada (HRDC) Longitudinal Labour Force File provides an example of failure in gauging support for data linking and reuse. HRDC was not an NSI but rather a government department, which had inherited both the remits and the data of other defunct government departments including: Labour Canada, Health and Welfare Canada, and the Employment Insurance Commission. From the records of these defunct departments HRDC created a database entitled the Longitudinal Labour Force File, which eventually contained around 2,000 pieces of information on each individual citizen. When attention was drawn to its existence by pressure groups and the media, the backlash was considerable. The database was termed a "citizen profile" by the Privacy Commissioner and became known as the "Big Brother Database" in the press.[61] The motivation for compiling this database was apparently neither certain nor particularly malicious.

How the linked-up records would be used and by whom, were questions that were not apparently clearly addressed. This meant that the product of data linking was vulnerable to "fishing expedition" searches of the database[62] that is to say there was no pre-specified rationale for access and reuse of the data. Indeed, Soleve may have been describing exactly this situation when he wrote the following words: "There is

---

[60] *Statistics and Registration Services Act 2007*, s. 39(2).

[61] Canadian Broadcasting Centre, Friday, November 10, 2000, at:

http://www.cbc.ca/canada/story/2000/05/29/bigbrother_000529.html.

[62] *Ibid*.

no diabolical motive or secret plan for domination; rather, there is a web of thoughtless decisions made by low-level bureaucrats, standardisation, policies, rigid routines and a way of relating to individuals and their information that often becomes indifferent to their welfare." [63] This situation was apparently one that the public, its representatives and indeed the Privacy Commissioner would not endorse. Clearly acknowledging the place of ethical as well as legal aspects of data access and reuse, the Privacy Commissioner responded in the following way to the HRDC's defence that it had not violated Canada's Privacy Act,

> *One does not have to be a privacy expert to see that this assertion rests on a restrictive and literal interpretation of the fundamental rights that are at the heart of the* Privacy Act. *I do not find it satisfactory that the federal government's largest department defends the creation, maintenance and expansion of dossiers on vast numbers of Canadians by saying that it meets minimum legal provisions…Surely a higher duty than that is imposed.*[64]

With respect to the lessons to be learned about the maintenance of public trust the Canadian case suggests that any matching or linking is done in as targeted a manner as possible. So for example, files could be linked together for the purpose of a particular piece of official statistical research and the link broken again. Canada's Privacy Commissioner used the analogy of silos and warehouses; [65] the former is a model of keeping data sources separate and protects privacy by default. "Not having a single client file is a good thing – on the principle that the more separate the databases the lower the risk of indiscriminate collection, unrelated uses and improper disclosure of personal data." [66]

8. Conclusion

Much has been done since the publication of the consultation paper *Statistics: a Matter of Trust* [67] to standardise existing practices across the decentralised system of official statistics in the UK. This has included the creation of framework documents for ONS and NS as well as the NS code of practice and the supporting protocols and now, most importantly, the Statistics and Registration Services Act. However, how far the Act directly addresses the ethical and social factors which have an impact on public trust is debatable. The act gives the UK's NSI a firm legal footing to access and reuse data which has been collected to fulfil administrative requirements. However, this may not go far enough in acknowledging existing political relations, especially for those producers of National Statistics based within the 'policy

---

[63] D Soleve, "Privacy and Power: Computer Databases and Metaphors for Information Privacy" (2001) 53 *Stanford Law Review* 1393-1462.

[64] Privacy Commissioner of Canada, Annual Report 1999-2000, 64-71, available at http://www.privcom.gc.ca.

[65] *Ibid*.

[66] *Ibid*.

[67] Office for National Statistics, "Statistics: A Matter of Trust" Report on the Consultation Exercise 24 February – 31 May 1998, Office for National Statistics, 1998.

departments'[68] producing the administrative data. As discussed above in relation to HRDC, even legally permissible activity may not be accepted by the general population. Moreover, achieving good practice and public support is arguably as dependent on the will of those involved in the production of official statistics as on the law itself. The *Statistics and Registration Services Act* does not permit the use of records for anything other than statistical or related purposes, it and the activities of the new Statistics Board operate in the wider context of the management and use of public records in the UK and the increasing interest of sections of the research community in reuse of administrative data. It is perhaps unfortunate, therefore that plans for a system for record linkage in the UK have tried to meet the goals of a wide variety of potential users. In a recent discussion paper on plans to create a joined up data source for the production of statistics ONS stated the following: "This database would underpin all ONS population and social statistics, resulting in significantly improved, more consistent statistics for the government community, the Health Service, academia and the private sector".[69] This may be interpreted by the public as a lack of clarity at policy level on the rationale for data linking. Moreover, anti-fraud and anti-terrorism measures following 11 September 2001, plus periodic imperatives to share information to avoid harm to the vulnerable in society following certain high profile cases[70], as well as the negligent loss of records and plans for ID cards, may create a generalised negative perception of public sector data management from which the Statistics Board are unlikely to benefit.

The path of linking records for reuse in statistical research is well worn in the Netherlands, while this largely new territory for the UK. The new Act is a step forward in that it changes the situation where government departments may have been acting *ultra vires* by data-sharing.[71] In addition, there are moves in the public sector to facilitate record linkage and reuse, such as NPfIT. However, the challenges may lay primarily with the social acceptance of a reuse of public administrative records. The issue of data being collected in such a way that makes it easy to reuse can only be partially solved by the introduction of new technologies and a legal framework. As Susan Leigh Star and James Griesemer point out, in order for an enterprise to be successful, in this case data linking and reuse, a variety of different groups must be willing to work together to share collective goals.[72] There are no clear provisions within the Act for stakeholder involvement in decisions around access and reuse of administrative data for statistical purposes. However, arguably, where the purposes of reuse are clear fears about privacy could be allayed; a well-defined approach may pre-empt the social and ethical objections potentially raised by the loss of "contextual integrity"[73] and ensure the support of the public and other stakeholders. Moreover, the argument that official statistics have the potential to do good for the wider community

---

[68] T Holt, *see* note 30.

[69] National Statistics, Proposals for an Integrated Population Statistics System, Discussion Paper Office for National Statistics, October 2003, Crown Copyright, 2003.

[70] *See* Bichard Inquiry, The Bichard Inquiry Report. London: Home Office, 2004, at: http://police.homeoffice.gov.uk/publications/operational-policing/bichard-inquiry-report.

[71] P Jackson, *see* note 18.

[72] S L Star & J R Griesemer, *see* note 37.

[73] H Nissenbaum, *see* note 57.

is likely to remain a powerful one for which people are willing to take additional risks. A strictly regulated, transparent and demarcated sphere for the linking and reuse of data could bring about a situation in which statisticians would have a good deal of freedom to access administrative records for statistical purposes. The lack of direct consent from data-subjects in the Netherlands for the supply of much of the data available to CBS is balanced against a regulatory framework that is both stringent and fairly transparent. As the Canadian case illustrates sustainable access to administrative records requires the support of interest groups, policy makers and the public. It is difficult to claim that the Dutch public and other stakeholders fully endorse all aspects of CBS's activities in this area. Indeed studies by CBS suggest that some citizens feel that data matching, even for statistical purposes, signifies a lack of individual control over how data is used[74]. However, CBS has maintained at least the tacit support of the public and of the different government departments which contribute data for statistical purposes, for several decades.

[74] J Nobel, *see* note 28.